A Clinical Database as a Component of a Diagnostic Hematology Workstation

Lawrence W. Diamond, Vladimir G. Mishka, Apollo H. Seal, and Doyen T. Nguyen
Pathology Institute, University of Cologne, Germany

A clinical database was designed as part of a comprehensive workstation for diagnostic laboratory hematology. The database stores coded findings pertinent to hematologic disorders including inherited abnormalities, previous surgery, malignancies, current therapy, and laboratory test results. The workstation includes knowledge-based systems for peripheral blood analysis, flow cytometry studies, and bone marrow morphology. The peripheral blood system renders an interpretive report based on data from a complete blood count with manual review of a blood smear by a technologist. The flow cytometry module interprets the immunophenotyping and DNA content results, and correlates them with the clinical findings and the peripheral blood data. The bone marrow system bases its report on all of the available information including the morphologic review of the bone marrow specimen by a physician, the peripheral blood data, immunophenotype, and clinical/laboratory findings. Before generating an interpretive report, each of the knowledge-based systems automatically searches the clinical database for specific information pertinent to the findings in the case. Since the workstation must function in situations where access to distributed databases is not feasible or not yet practical, a data entry module with a graphical user interface has been created.

INTRODUCTION

Knowledge-based systems for decision support in medical diagnosis and therapy need access to the same clinical and laboratory data that a physician would use in working-up a patient. If the decision support modules are part of a hospital information system interfaced to a laboratory computer as in the HELP system [1], access to existing on-line information is facilitated. The extensive work being done on electronic patient records [2] and open architectures for distributed databases [3] will ultimately improve access to the needed clinical information. In the meantime, there is still a role for local databases integrated into workstations designed for interpretive reporting, education, and clinical research.

We have designed a hematology workstation [4] with knowledge-based systems for peripheral blood interpretation [5], flow cytometry (FCM) immunophenotyping and DNA content analysis [6], and bone marrow morphology [7]. The three modules communicate with each other by means of a set of relational databases [8]. Each of the systems modifies its report based on the interpretation stored in the respective databases by the other modules. The patient database has been designed to include coded information relevant to hematologic disorders from the clinical history, physical examination, radiology reports, and laboratory tests outside of hematology. In this paper, we describe the design of this database and the role it plays in the hematology workstation.

KNOWLEDGE-BASED SYSTEMS

The workstation is interfaced to state-of-the-art hematology analyzers such as the Coulter STKS and ONYX. The STKS is a high performance analyzer with five-part automated leukocyte differentials suitable for processing several hundred to several thousand complete blood counts (CBCs) each day. The ONYX has automated three-part differentials and is intended for mid-range laboratories with a moderate volume. The workstation is being tested on two hardware platforms, one for STKS sites and one for ONYX laboratories. The computer at the STKS test sites is an NEC Image 466es (486-DX2 66 MHz) with 20 MB of RAM and a 540 MB hard disk drive running Windows NT (Microsoft Corporation, Redmond, WA). The ONYX sites are using an Intel 486sx-25MHz computer with 8 MB of RAM and a 170 MB hard disk drive running Windows 3.1. The knowledge-based systems have been programmed with Turbo Pascal for Windows,

0195-4210/94/$5.00

version 1.5, and the database components have been designed with Paradox Engine, version 3.01 (Borland International, Scotts Valley, CA). All knowledge engineering and computer programming have been carried out by a trained hematopathologist.

The three knowledge-based systems have been described in detail previously [4-8]. Briefly, the peripheral blood system, known as Professor Petrushka, interprets the hemogram data downloaded across the interface from the hematology analyzer [5]. In order to maintain just one version of Petrushka, the database components for peripheral blood specimen results must be the same for both the STKS systems (which download five-part leukocyte differentials and complex three-dimensional scatterplots) and the ONYX systems (which download three-part differentials and simple one-parameter histograms). The databases are identical for the two systems with the exception of the names of the database fields which hold the graphics (binary large object fields). At startup, Petrushka checks the name of one of the graphics fields in the database and automatically configures itself to display the appropriate screens for a STKS or ONYX system. Professor Petrushka determines the predominant pattern of the hemogram parameters, and, if the preparation of a peripheral blood smear is indicated, recommends a specific approach to the review of the smear. A technologist enters red blood cell morphology with a mouse. The workstation keyboard also functions as a differential counter for doing manual leukocyte differentials. Petrushka's final interpretation takes into account the data entered by the user, pertinent clinical information and laboratory test results retrieved from the clinical database, and the results of previous CBC specimens. The interpretive reports include the predominant pattern of the peripheral blood data, a differential diagnosis, any features which indicate a more specific diagnosis, and recommendations for additional clinical history and laboratory tests which may be helpful in making a final diagnosis. If, for example, a reticulocyte count is indicated, the database is checked for reticulocyte results. If none are found, Petrushka recommends an absolute reticulocyte count. If the results are already posted to the database, Petrushka interprets the reticulocyte count in light of the rest of the CBC findings.

The FCM system, Professor Fidelio, interprets the immunologic data in leukemias/lymphomas from a panel of up to 37 antibodies, and DNA content data including cell size by light scatter, DNA index (a measure of DNA ploidy) and the percentage of cells in the S-phase of the cell cycle [6]. The system is also used for T-cell subset analysis and follow-up in AIDS patients. Patient-specific clinical data stored in the database, such as a previous history of chronic myeloid leukemia, are used to modify the interpretation of the results. For example, immunologic findings which are otherwise compatible with a diagnosis of acute lymphoid leukemia, are indicative of a blast crisis of chronic myeloid leukemia when the Philadelphia chromosome was documented by previous cytogenetic studies.

When a patient identification number is entered into Professor Belmonte [7], the module for bone marrow morphology, the program searches the peripheral blood specimen database for any recent results and uses them to remind the physician of any additional studies to be performed on the bone marrow specimen, such as FCM immunophenotyping and cytogenetics. Using Belmonte's graphical user interface, a physician enters the qualitative morphologic findings and the bone marrow differential. The program then summarizes the peripheral blood findings and checks the FCM and patient databases for specific information before writing an interpretive report. For example, in a patient with acute leukemia, the diagnosis of an acute myelomonocytic leukemia (AML-M4) requires either a positive stain for non-specific esterase (a monocyte marker) in bone marrow blasts or an increased lysozyme in the serum or urine. If the non-specific esterase stain was negative or not done, and the bone marrow findings are otherwise suggestive of AML-M4, Belmonte will check the database for the lysozyme results in order to establish the diagnosis. If, in this example, no lysozyme results are available, the system will not subtype the acute myeloid leukemia but will append a comment to the report indicating the need for this test.

DATABASE DESIGN

The patient database for clinical findings and laboratory test results has been specifically designed for speed and integration into the hematology workstation. It is intended to

capture only those findings which are appropriate to diagnostic hematology. No attempt has been made to capture all of the history, physical examination findings, radiology results, or laboratory test results appropriate for all of internal medicine. The fields in the patient database are listed in Table 1. The last nine fields listed in the table contain the coded information for specific pieces of clinical and laboratory data. The information is coded in strings of characters. Fields 7-14 each contain space for 16 alphanumeric characters. The last field, laboratory tests, contains 24 alphanumeric characters. Unused spaces at the end of the strings have been reserved for future expansion, when other data relevant to hematologic diseases are added to the medical repertoire.

Table 1
Patient Database Fields

1. Patient Identification Number
2. Name
3. Sex
4. Date of Birth
5. Date of Data Entry
6. Race
7. Inherited diseases
8. Social History
9. Previous Surgery
10. Current Drug Administration
11. Cancer History
12. Autoimmune Diseases
13. Physical Examination/Radiology
14. Laboratory Tests for Impaired Immunity
15. General Laboratory Tests

The medical data is coded into the strings by allowing each character in the string to represent one piece of information. Unknown or missing data is represented by a blank space. A negative finding is represented by a '0' (zero). Positive findings are represented by other ASCII characters. For example, lymphoproliferative disorders are coded in the fourth character of the cancer history field. The codes for lymphoproliferative disorders and their meaning are listed in Table 2.

Using this scheme, retrieval of various types of information is very efficient. For instance, Professor Petrushka may need to know if a patient is currently on chemotherapy as an explanation for pancytopenia. If this information is missing, the interpretive report may contain a comment to check for this possibility. In this case, the system needs only to retrieve the drug history string using the patient's identification (medical record) number as the key field, and check if the twelfth character is "less than" zero, equal to zero or "greater than" zero. If it is less than zero, then it must be a blank space which means that the information is missing and the comment should be appended. If the character is zero, then the patient is not on chemotherapy and a different explanation must account for the pancytopenia. The finding that the character is greater than zero confirms chemotherapy as the cause of the pancytopenia. The combination of chemotherapeutic agents being administered is contained in the value of the character, but Petrushka will not bother to decode that information unless it is necessary to the generation of the interpretive report.

Table 2
Codes for Lymphoproliferative Disorders

'Blank'	Missing Data
0.	No evidence of disease
1.	Lymphoproliferative Disorder, NOS
2.	Chronic Lymphocytic Leukemia
3.	Prolymphocytic Leukemia
4.	Hairy Cell Leukemia
5.	Multiple Myeloma
6.	Waldenstrom's Macroglobulinemia
7.	Adult T-cell Leukemia/Lymphoma
8.	Large Granular Lymphocytosis
9.	Mycosis Fungoides/Sezary Syndrome
A.	Plasmacytoma

NOS: Not otherwise specified

The value of many types of medical information is time dependent. A patient who has just had a bone marrow transplant is an example. The critical information needed to interpret CBC and bone marrow findings properly depends on the length of time since the transplant was performed. The length of time post transplant is coded in six intervals ranging from less than one week to greater than two months. The knowledge-based systems (Petrushka and Belmonte) check the date of the

clinical data and compare it to the date of the specimen, re-calculating the information contained in the database, if necessary. If a patient was one week post transplant when the patient database was last updated, and a week has elapsed between the date stamped in the database entry and the date of the specimen, the programs will interpret the findings in light of the knowledge that the patient is now two weeks post transplant.

In addition, in this scheme of data encoding, unused characters can be reserved to keep track of time. For example, in a patient who has been treated with combination chemotherapy and radiotherapy for Hodgkin's disease, and who subsequently develops an acute myeloid leukemia, it is important to know the interval of time between the start of therapy and the onset of leukemia, in order to determine if the leukemia is therapy related. The time factor can be stored in the database by a cumulative counter which keeps track of the number of days since the start of therapy (in hexadecimal notation) in four characters of the string. The counter is started at '0000' at the time when current chemotherapy is noted in the patient record. The database program adds to the counter the number of days which have elapsed when the record is subsequently updated. This same scheme can be used to time laboratory results and 'discard' them from influencing the interpretation after they have 'expired'.

In order to allow for the possibility of manual entry of clinical data, a Windows program with a graphical user interface was developed to store the coded information (Figure 1). With the exception of the medical record number, no typing is required to enter data.

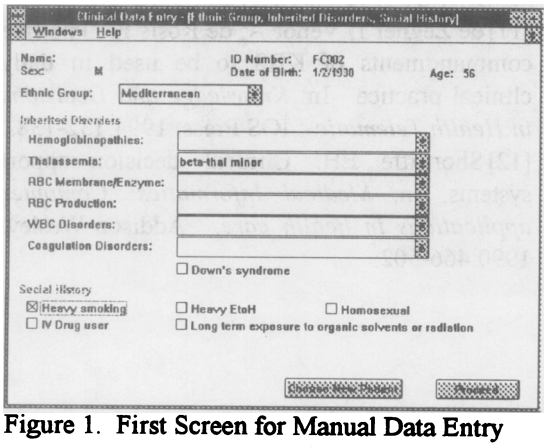

Figure 1. First Screen for Manual Data Entry

DISCUSSION

In the design of a comprehensive workstation for diagnostic laboratory hematology, we have decided to keep track of relevant patient-specific clinical and laboratory information. The power of an intelligent computer system lies in its knowledge base. Therefore, the design of our clinical database has been influenced by our knowledge of the practice of laboratory hematology. At present, we have limited ourselves to a single patient clinical record. Although this may be viewed as a weakness by some database developers, much of the information we keep track of, such as inherited diseases, malignancies, or surgery are effectively mapped to a patient in a single record. As has been described above, we have also taken into consideration methods which will allow us to keep track of time for specific items within the context of our database design.

Our approach is best suited to diagnosis. We acknowledge that systems designed primarily for supporting therapy, like ONCOCIN, are likely to need a different database design based on a time-oriented model of reasoning, with multiple records for the same patient [9].

Miller and Maserie have utilized an elegant and efficient algorithm for the QMR relationships function by taking advantage of compiled bitmapped data structures [10]. Our clinical database strings can be thought of as "byte-mapped" structures which allow for fast and efficient use of positive, negative, or missing clinical information by our knowledge-based systems.

Although our clinical database is specific to our applications, the principles used in the design may be useful to other researchers developing real-world diagnostic systems which must be capable of existing in both highly networked hospitals (with access to distributed databases), and small satellite facilities without current access to electronic medical records.

Two of the most important issues that have been identified as necessary features for a knowledge-based system to function beyond the author's laboratory and really be used in medical practice are: (1) Working with patient data contained in the medical record; and (2) Integration into the daily routine of medical practice [11]. Our knowledge-based systems are designed with both of these goals in mind. Our

clinical database represents a medical record of hematology data which is scanned by each of our modules as they generate an interpretive report. The systems modify the report to reflect the clinical and laboratory data. In addition, when a case is interpreted by the bone marrow system, important diagnoses are posted back to the database to keep the record up-to-date.

For integration into the hematology laboratory environment, our workstation is designed to be used on-line for interpretive reporting and clinical research in addition to education. We have carefully followed the established protocols for system development with respect to design, prototyping, human-computer interface issues, and in-house validation of the knowledge base [12].

The workstation has now progressed to the phase of large scale outside evaluation at many laboratories in multiple countries. The first outside evaluations took place at the University of Florida and M.D. Anderson Cancer Center in late 1993. The workstation was installed (April 1994) as an ONYX site at Scott and White Hospital and Clinics, Temple Texas, a tertiary medical center with 18 satellite laboratories. In June 1994, systems were installed at both the University of Cincinnati, and MDS Laboratories in Toronto, Canada (a reference laboratory processing 3000 CBCs per day). Three sites in the United Kingdom, Addenbrookes Hospital, (Cambridge), St. Thomas Hospital (London), and St. James Hospital (Dublin) have agreed to become test sites with installations planned at the end of 1994.

At the large medical centers, the next phase of development will be interfaces/networking to existing hospital and laboratory information systems for automatic retrieval and processing of the information which is stored in our patient database. In addition, at MDS laboratories, work is underway to integrate the workstation into their state-of-the-art computer-controlled laboratory robotics system.

Each of the test sites has been impressed with the initial performance of the workstation. Currently, the manual data entry module is being used to enter clinical information. As clinical data is collected and reviewed, it is used to increase the number of clinical scenarios recognized by the knowledge-based systems.

References

[1] Nelson BD, Gardner RM. Decision support for concurrent utilization review using a HELP-embedded expert system. *SCAMC*, 1993;17:176-182.

[2] Rind DM, Safran C. Real and imagined barriers to an electronic medical record. *SCAMC*, 1993;17:74-78.

[3] Chueh HC, Barnett GO. Client-server distributed database strategies in a healthcare record system for a homeless population. *SCAMC*, 1993:17:119-124.

[4] Nguyen DT, Diamond LW. The concept of a computer workstation for laboratory haematology. *Aust J Med Sci*, 1993;14:71-75.

[5] Nguyen DT, Diamond LW, Priolet G, Sultan C. A decision support system for diagnostic consultation in laboratory hematology. *MEDINFO*, 1992;7:591-595.

[6] Diamond LW, Nguyen DT, Jouault H, Imbert M. Evaluation of a knowledge-based system for interpreting flow cytometric immunophenotyping data. *MIE*, 1993;11:124-128.

[7] Nguyen DT, Cherubino P, Tamino PB, Diamond LW. Computer-assisted bone marrow interpretation: A pattern approach. *MIE*, 1993;11:119-123.

[8] Diamond LW, Nguyen DT. Communication between expert systems in haematology. In: *Current Perspectives in Healthcare Computing*: BJHC Books, 1993:111-119.

[9] Kahn MG, Fagan LM, Tu S. Extensions to the time-oriented database model to support temporal reasoning in medical expert systems. *Meth Inform Med*, 1991;30:4-14.

[10] Miller RA, Maserie FE Jr. The Quick Medical Reference (QMR) relationships function: Description and evaluation of a simple efficient multiple diagnoses algorithm. *MEDINFO*, 1992;7:512-518.

[11] de Zegher I, Venot A, de Rosis F. The five commandments of KBS to be used in daily clinical practice. In: *Knowledge and Decisions in Health Telematics*: IOS Press, 1994:132-138.

[12] Shortliffe EH. Clinical decision-support systems. In: *Medical Informatics: Computer applications in health care*: Addison-Wesley, 1990:466-502.